INCORPORATING MULTI-CHANNEL WIENER FILTER WITH SINGLE-CHANNEL SPEECH ENHANCEMENT ALGORITHM

Pei Chee Yong, Sven Nordholm, Hai Huyen Dam, Yee Hong Leung, Chiong Ching Lai

Curtin University, Kent Street, Bentley, WA 6102, Australia
Peichee.Yong@postgrad.curtin.edu.au {S.Nordholm, H.Dam, Y.Leung, Chiong.Lai}@curtin.edu.au

ABSTRACT

The real-time implementation of the existing multi-channel Wiener filter (MWF) algorithms suffer from performance degradation due to the lack of robustness against estimation errors of the second-order statistics. The reasons are twofold: one, the estimation of the statistics relies on real voice activity detector (VAD), which often fails in adverse environments. Second, the MWF solutions involve estimation of the second order clean speech statistics, which also exaggerates the errors. This paper presents an MWF algorithm that requires neither VAD nor clean speech statistics. Performance evaluation under real scenarios shows that the proposed method outperforms the conventional MWF solution in terms of the trade-off between noise reduction and speech distortion.

Index Terms— Multi-channel Wiener filter, speech enhancement, single-channel noise reduction

1. INTRODUCTION

Research in speech enhancement has been active for many years due to its diverse applications ranging from telecommunication devices to assistive listening devices. Among the multi-channel techniques reported in the literature, speech distortion weighted multi-channel Wiener filter (SDW-MWF) [1, 2] is promising as it does not require prior knowledge about the location of the desired speech signal and the microphone characteristics. As a result, it is more robust against microphone mismatch when compared to the well-known beamformer, the generalized sidelobe canceller (GSC) [3]. Similar to the GSC, SDW-MWF relies on a voice activity detection (VAD) algorithm to update the noise statistics in noise-only segments, and the signal statistics during voiced segments. As a VAD estimate is required in practice, wrong estimation often occurs under non-stationary and highly noisy environments, which leads to greater second order estimation errors and causes performance degradation in the SDW-MWF method [4, 5].

Alternatively, the SDW-MWF solution can be decomposed into a rank-one problem, namely the R1-MWF method that consists of a spatial filter and a single-channel postfilter [6, 7]. Although R1-MWF is more robust against the estimation errors, the single-channel postfilter may not be optimal in terms of spectral tracking, since it is based on correlation matrices that are adapted slowly over time. This has been improved by using a multi-channel speech presence probability (MC-SPP) algorithm to adapt the noise statistics continuously over time [8]. Instead of using MC-SPP, a more direct SPP estimate can be obtained by taking one of the microphone inputs as reference, which has been used in [9] to adapt the parameter that trades off noise reduction and speech distortion. To increase the accuracy in speech detection, both MC-SPP and SPP require accurate estimates of *a priori* speech absence probability (SAP) and *a priori* signal-to-noise ratio (SNR), which in turn increase the processing delay. To avoid this, fixed prior estimates can be used not only to reduce the delay but also to maintain the accuracy in noise tracking in single-channel speech enhancement framework [10, 11].

This paper proposes a non-VAD based SDW-MWF solution that aims to reduce the second order statistics errors by avoiding the subtraction of noise-only correlation matrix from the speech-plus-noise correlation matrix. In that case, the desired signal is estimated from the reference microphone by using a single-channel speech enhancement framework from [12], which shows good performance in terms of trade-off between noise reduction, speech distortion and musical noise. In addition, the noise power spectral density (PSD) estimate in the reference channel is obtained by the modified SPP with fixed priors approach in [11], which is then employed to continuously update both noisy and noise second order statistics.

The paper is organized as follows. Section 2 shows the conventional MWF solutions. Section 3 develops the proposed methods. Section 4 presents the results and Section 5 concludes the paper. Section 6 shows the relation between the contribution in this paper and prior works in the field.

2. MULTI-CHANNEL WIENER FILTER

2.1. Signal model and notation

Let $Y_l(k,m)$, $l = 1, ..., L$, denote the microphone signals in time-frequency domain, where k is the frequency bin index,

This research was supported in part by Sensear Pty Ltd and Linkage Grant LP100100433 by Australian Research Council (ARC).

m is the frame index and L is the number of microphones. The received signals are given by

$$Y_l(k,m) = X_l(k,m) + V_l(k,m) \tag{1}$$

where $X_l(k,m)$ and $V_l(k,m)$ are the short-time Fourier transform (STFT) representations of the target signal and the uncorrelated noise components, respectively, of the l^{th} microphone. Here, speech enhancement is performed to remove the unwanted noise while preserving the target speech signal. This can be done by applying a set of filter $\mathbf{w}(k,m)$ to the observed signal, such that

$$Z(k,m) = \mathbf{w}^H(k,m)\,\mathbf{y}(k,m) \tag{2}$$

where Z is the output signal, and $\mathbf{y}(k,m) \in \mathbb{C}^{L\times 1}$ is a stacked vector given as

$$\begin{aligned}\mathbf{y}(k,m) &= [Y_1(k,m)\ Y_2(k,m)\ ,...,\ Y_L(k,m)]^T \\ &= \mathbf{x}(k,m) + \mathbf{v}(k,m)\end{aligned} \tag{3}$$

with T indicating the transpose operator. From now on, both indices (k,m) will be omitted for notational convenience. The correlation matrices for the noisy speech $\mathbf{R}_y$, the clean speech $\mathbf{R}_x$, and the background noise $\mathbf{R}_v$ are then defined, respectively, as

$$\mathbf{R}_y = E\left\{\mathbf{y}\mathbf{y}^H\right\},\ \mathbf{R}_x = E\left\{\mathbf{x}\mathbf{x}^H\right\},\ \mathbf{R}_v = E\left\{\mathbf{v}\mathbf{v}^H\right\}, \tag{4}$$

where E and H denote, respectively, the expected value and Hermitian transpose operators.

2.2. Formulation of multi-channel Wiener filter

The multi-channel Wiener filter (MWF) optimally estimates the speech signal, based on an MMSE criterion, as

$$\mathbf{w}_{\text{MWF}} = \arg\min_{\mathbf{w}}\ E\left\{\left|X_{\text{ref}} - \mathbf{w}^H\mathbf{y}\right|^2\right\} \tag{5}$$

where the desired signal in this case is the unknown speech component X_{ref} from the reference microphone. The drawback is that some residual noise will still remain in the output signal, Z, which can be reduced by allowing a trade-off between noise reduction and speech distortion. This can be done by modifying the design criterion of the MWF as [1, 6]

$$\mathbf{w}_{\text{MWF}_\mu} = \arg\min_{\mathbf{w}}\ E\left\{\left|X_{\text{ref}} - \mathbf{w}^H\mathbf{x}\right|^2\right\} + \mu E\left\{\left|\mathbf{w}^H\mathbf{v}\right|^2\right\} \tag{6}$$

where speech and noise are assumed to be uncorrelated, and μ is the trade-off parameter. A larger μ value here indicates more residual noise reduction at the expense of higher speech distortion. The solution of MWF_μ can then be obtained as

$$\mathbf{w}_{\text{MWF}_\mu} = \left[\mathbf{R}_x + \mu\mathbf{R}_v\right]^{-1}\mathbf{R}_x\mathbf{e}_{\text{ref}} \tag{7}$$

where $\mathbf{e}_{\text{ref}} = [0...0\ 1\ 0...0]^T$ is a L-element zero vector with the unity corresponds to the r^{th} element of the microphones. Here, the correlation matrices $\mathbf{R}_y$ and $\mathbf{R}_v$ can be recursively updated by using a VAD, as

$$\begin{aligned}\mathcal{H}_0 &: \begin{cases}\hat{\mathbf{R}}_v[m] = (1-\alpha_v)\,\hat{\mathbf{R}}_v[m-1] + \alpha_v\mathbf{y}[m]\,\mathbf{y}^H[m] \\ \hat{\mathbf{R}}_y[m] = \hat{\mathbf{R}}_y[m-1]\end{cases} \\ \mathcal{H}_1 &: \begin{cases}\hat{\mathbf{R}}_y[m] = (1-\alpha_y)\,\hat{\mathbf{R}}_y[m-1] + \alpha_y\mathbf{y}[m]\,\mathbf{y}^H[m] \\ \hat{\mathbf{R}}_v[m] = \hat{\mathbf{R}}_v[m-1]\end{cases}\end{aligned} \tag{8}$$

where $\mathcal{H}_0$ and $\mathcal{H}_1$ denote speech absence and speech presence in the k^{th} frequency bin of the m^{th} frame, respectively. Both smoothing factors α_y and α_v have to be chosen carefully to reflect the degree of stationarity of speech and noise signals.

From Eq. (7), it can be observed that an estimation of $\mathbf{R}_x$ is required, which is usually obtained by $\mathbf{R}_y - \mathbf{R}_v$ [1]. However, estimation errors in both the complex-valued correlation matrices $\mathbf{R}_y$ and $\mathbf{R}_v$ can result in a very poor estimate of $\mathbf{R}_x$. Although this can be avoided by obtaining a pre-determined $\mathbf{R}_x$ estimate either with a calibration sequence [13], or by deriving a mathematical model [14, 15], these methods rely on the *a priori* information, which makes them less attractive for on-line applications.

3. PROPOSED METHOD

3.1. Formulation of MWF_λ and estimation of noisy and noise correlation matrices

In order to avoid the aforementioned problems, a bi-criteria optimization problem for MWF is proposed, which consists of a criterion to minimize the error in Eq. (5) and another criterion to minimize the noise power. One way to formulate such problem is to use the weighted sum between the two criteria as given by

$$\begin{aligned}\mathbf{w}_{\text{MWF}_\lambda} = \arg\min_{\mathbf{w}}\quad & (1-\lambda)\,E\left\{\left|X_{\text{ref}} - \mathbf{w}^H\mathbf{y}\right|^2\right\} + \\ & \lambda\left(E\left\{\left|\mathbf{w}^H\mathbf{v}\right|^2\right\}\right)\end{aligned} \tag{9}$$

where λ is a weighting value between 0 and 1. The solution of the problem can then be found as

$$\mathbf{w}_{\text{MWF}_\lambda} = \left[(1-\lambda)\,\mathbf{R}_y + \lambda\mathbf{R}_v\right]^{-1}(1-\lambda)\,\mathbf{r}_{yx} \tag{10}$$

where $\mathbf{r}_{yx} = E\{\mathbf{y}X_{\text{ref}}^*\}$. It can be seen that by formulating the problem in this way, the estimation of the clean speech correlation matrix $\mathbf{R}_x$ can be averted. Also, a set of pareto solutions can be found by varying λ, but this is not in the scope of this paper.

Apart from that, instead of using a VAD to estimate the correlation matrices, the frame and frequency dependant modified SPP, p from [11] is employed, which allows both $\hat{\mathbf{R}}_v$ and $\hat{\mathbf{R}}_y$ from Eq. (8) to be updated as

$$\hat{\mathbf{R}}_v[m] = (1-\tilde{\alpha}_v[m])\,\hat{\mathbf{R}}_v[m-1] + \tilde{\alpha}_v[m]\,\mathbf{y}[m]\,\mathbf{y}^H[m] \tag{11}$$

$$\hat{\mathbf{R}}_y[m] = (1 - \tilde{\alpha}_y[m])\,\hat{\mathbf{R}}_y[m-1] + \tilde{\alpha}_y[m]\,\mathbf{y}[m]\,\mathbf{y}^H[m] \quad (12)$$

where $\tilde{\alpha}_v$ and $\tilde{\alpha}_y$ are given respectively by $\tilde{\alpha}_v = \alpha_v(1-p)$ and $\tilde{\alpha}_y = \alpha_y p$, with α_v and α_y denote, respectively, the fixed smoothing factor for noise and noisy correlation matrices.

3.2. Employing single-channel algorithm

From Eq. (10), it can be seen that the proposed solution requires an estimate of the clean speech reference X_{ref}. As opposed to previous methods [13, 14, 15], we propose to estimate X_{ref} by utilizing a single-channel speech enhancement method and to use one microphone in the array as a reference. As such, the estimate of $\mathbf{r}_{yx}$ can be defined as

$$\hat{\mathbf{r}}_{yx} = \mathbf{y}G(X^*_{\text{ref}} + V^*_r) \quad (13)$$

where $X_{\text{ref}} = H_{\text{ref}}S$ with H_{ref} denotes the acoustic transfer function of the target speech signal, S at the reference channel. Here, G is a spectral weighting gain function, which involves the computation of the *a posteriori* and *a priori* SNR estimates. In contrast to $\mathbf{R}_x\mathbf{e}_{\text{ref}} = (\mathbf{R}_y - \mathbf{R}_v)\mathbf{e}_{\text{ref}}$ from Eq. (7), which takes the reference vector directly from the second order clean speech estimate, Eq. (13) uses an SNR based gain function to adapt the noisy stacked vectors to the desired clean speech signal. Such implementation is capable of generating a better clean speech estimate and improving the speech quality of the enhanced signal.

In this paper, G in Eq. (13) is taken from the modified sigmoid (MSIG) gain function from [12]. As the beamformer tries to adapt to the clean speech reference, an important aspect of the single-channel estimate is that the speech distortion has to be as small as possible. This can be done by setting smaller values for the SNR smoothing parameters from [12], i.e. $\beta \approx 0.9$ and $\alpha_y \approx 0$, such that the amount of speech distortion can be kept as low as possible while not having a large amount of musical noise. Apart from that, further reduction of musical noise is proposed by having $\hat{\mathbf{r}}_{yx}$ updated recursively as

$$\hat{\mathbf{r}}_{yx}(m) = (1-\alpha_x)\hat{\mathbf{r}}_{yx}(m-1) + \alpha_x\mathbf{y}(m)\,\hat{X}^*_{\text{ref}}(m) \quad (14)$$

where α_x is the smoothing factor for target speech signal, and $\hat{X}_{\text{ref}} = G(X_{\text{ref}} + V_r)$ indicates the clean speech estimate from the reference microphone.

The SNR estimates for G require the estimation of the noise PSD at the reference channel. Here, the noise PSD estimate in [11] is used, which involves the calculation of the modified SPP. This implies that the same SPP estimate can be used for estimating the noise PSD in the reference channel and also the correlation matrices in Eqs. (11) and (12).

4. PERFORMANCE EVALUATION

Measurements are performed with 2 microphones (with inter-element space of 1 cm) embedded in the left side of a pair of earmuffs on a manikin so that the head-shadowing effect is included. The manikin is placed close to the center of a room with dimensions $3.05\,\text{m} \times 3.05\,\text{m}$, with a reverberation time T_{60} of approximately $0.2\,\text{s}$. The loudspeakers are positioned at 1 m from the center of the head, with the speech located at 0° and the non-stationary factory noise rendered at 45°, 90°, 135°, 180°, 225°, 270° and 315° to the left of the head. The speech signals consists of 5 (2 male and 3 female) sentences with length ranging from 11 s to 22 s. The signals are sampled at $f_s = 16\,\text{kHz}$. An STFT length of $K = 512$ is used with a frame rate $R = 256$ and square-root Hann windowing.

Evaluation includes $\mathbf{w}_{\text{MWF}_\mu}$ from Eq. (6) with $\mu = 5$, the output signal from reference microphone using MSIG function with a noise floor of $-15\,\text{dB}$, $\mathbf{w}_{\text{MWF}_{\lambda 1}}$ from Eq. (10) with $\lambda \approx (\mu-1)/\mu = 0.8$ and $\mathbf{w}_{\text{MWF}_{\lambda 2}}$ with $\lambda = 1-p$. The smoothing constants are estimated by $\alpha = \exp(\frac{-2.2R}{tf_s})$, with $t_x = t_y = 0.02\,\text{s}$ and $t_v = 2\,\text{s}$. The performance is measured by the speech intelligibility weighted segmental SNR in frequency domain (IFWSNRseg) [16, 17]

$$\text{IFWSNRseg} = \frac{10}{M}\sum_{m=0}^{M-1}\frac{\sum_{k=0}^{K-1} B_k \log_{10}\frac{A^2(k,m)}{A^2(k,m)-\hat{A}^2(k,m)}}{\sum_{k=0}^{K-1} B_k} \quad (15)$$

where B_k is the ANSI SII weight placed on the k^{th} frequency bin [18], K is the number of bands, M is the number of frames, $A(k,m)$ and $\hat{A}(k,m)$ are spectrum amplitudes of the clean speech signal and enhanced speech signal, respectively. Each frame is threshold by a $-10\,\text{dB}$ lower bound and a $35\,\text{dB}$ upper bound to discard non-speech frames.

In addition, noise attenuation (NATTseg) and speech preservation (SPREseg) measures are utilized to study if a difference in IFWSNRseg is due to more noise reduction or less speech distortion. Both are given, respectively, by [19]

$$\text{NATTseg} = \frac{1}{M}\sum_{m=0}^{M-1} 10\log_{10}\frac{||\mathbf{v}_t(m)||^2}{||\mathbf{G}(m)\mathbf{v}_t(m)||^2} \quad (16)$$

$$\text{SPREseg} = \frac{1}{M}\sum_{m=0}^{M-1} 10\log_{10}\frac{||\mathbf{x}_t(m)||^2}{||\mathbf{x}_t(m)-\mathbf{G}(m)\mathbf{x}_t(m)||^2} \quad (17)$$

where $\mathbf{v}_t(m)$ and $\mathbf{x}_t(m)$ are m^{th} frame time-domain vectors for the noise and the clean speech signal, respectively. The filtering matrix $\mathbf{G}$ indicates that both the noise and the clean signals are processed with the same corresponding filters as used to enhance the noisy signal. Apart from that, the widely-used perceptual evaluation of speech quality (PESQ) measure has also been included for performance comparison [17].

Figs. 1-4 show the average results for SNRs of $-5\,\text{dB}$, $0\,\text{dB}$, $5\,\text{dB}$, and $10\,\text{dB}$, respectively. It can be observed that $\mathbf{w}_{\text{MWF}_{\lambda 1}}$ outperforms $\mathbf{w}_{\text{MWF}_\mu}$ for all objective measures in all scenarios, indicating that the proposed method allows more noise suppression, which does not come with higher speech distortion. When $\mathbf{w}_{\text{MWF}_{\lambda 1}}$ is compared to MSIG and

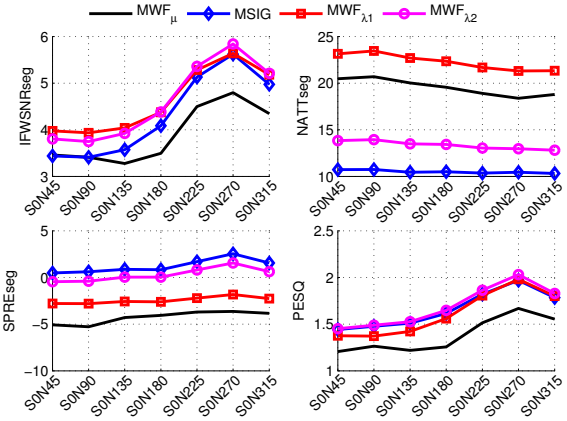

Fig. 1. Average results for input SNR -5 dB

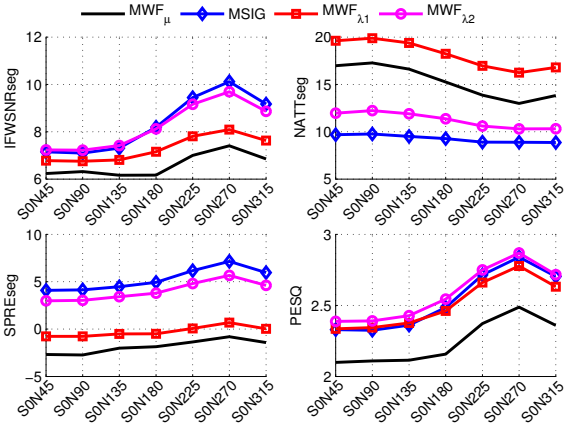

Fig. 3. Average results for input SNR 5 dB

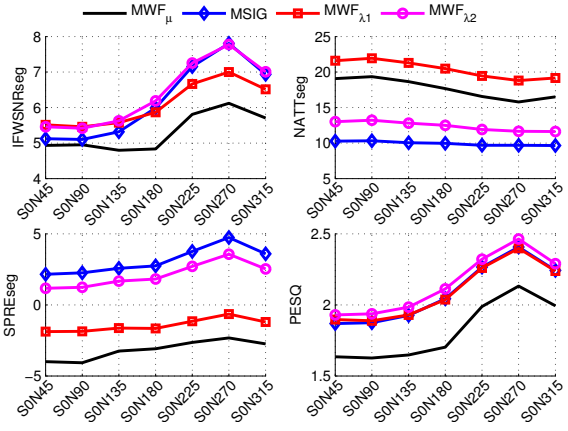

Fig. 2. Average results for input SNR 0 dB

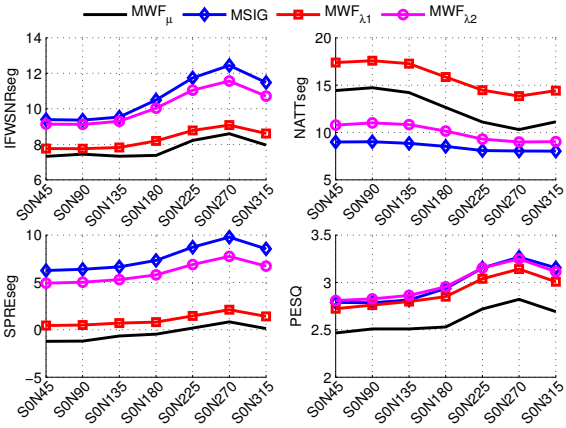

Fig. 4. Average results for input SNR 10 dB

$\mathbf{w}_{\mathrm{MWF}_{\lambda2}}$, it generally has larger noise reduction but larger speech distortion as well. This is the reason why $\mathbf{w}_{\mathrm{MWF}_{\lambda1}}$ performs better at low input SNR conditions but having performance drop when the input SNR increases, relatively to other evaluated methods, as shown in IFWSNRseg and PESQ results. While $\mathbf{w}_{\mathrm{MWF}_{\lambda2}}$ improves the performance of $\mathbf{w}_{\mathrm{MWF}_{\lambda1}}$ in terms of less speech distortion, more musical noise is audible since the NATTseg values are much lower than $\mathbf{w}_{\mathrm{MWF}_{\lambda1}}$. When compared to MSIG from the reference microphone, $\mathbf{w}_{\mathrm{MWF}_{\lambda2}}$ has higher noise reduction but also larger speech distortion. However, since MWF involves temporal averaging in the second order statistics estimation, musical noise can be reduced, especially at low SNR conditions, as indicated by IFWSNRseg results from Fig. 1.

5. CONCLUSIONS

In this paper, the proposed formulation of SDW-MWF has avoided the estimation of second-order clean speech statistics by incorporating a single-channel speech enhancement framework in estimating the desired signal from a reference microphone. Experimental results show that the proposed method outperforms the traditional method for all performance measures. By incorporating SPP in the trade-off parameter λ helps to reduce speech distortion, but in turn generates more residual noise and musical noise in the enhanced signals.

6. RELATION TO PRIOR WORK

This paper has focused on an alternative SDW-MWF formulation that does not require the clean speech correlation matrix estimate, as opposed to previous formulations in the literature [1, 6, 7]. Furthermore, as in contrast to work that requires calibration [13] or pre-calculation using a mathematical model [14, 15], this work utilizes single-channel noise reduction technique to estimate a reference channel, which as far as we are aware it has not been considered before. In addition, unlike the previous approach, where SPP was only used to adapt the trade-off parameter [9], or only to estimate the noise correlation matrix [8], it is fully utilized by the proposed framework in estimating the noise PSD in the reference channel and also both noisy and noise correlation matrices.

7. REFERENCES

[1] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction," *Speech Communication*, vol. 49, no. 7, pp. 636–656, Jul. 2007.

[2] A. Spriet, M. Moonen, and J. Wouters, "Stochastic gradient-based implementation of spatially preprocessed speech distortion weighted multichannel Wiener filtering for noise reduction in hearing aids," *IEEE Trans. on Signal Process.*, vol. 53, no. 3, pp. 911–925, Mar. 2005.

[3] L. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. on Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, Jan. 1982.

[4] A. Spriet, M. Moonen, and J. Wouters, "Robustness analysis of multichannel Wiener filtering and generalized sidelobe cancellation for multimicrophone noise reduction in hearing aid applications," *IEEE Trans. on Speech and Audio Process.*, vol. 13, no. 4, pp. 487–503, Jul. 2005.

[5] S. Doclo, M. Moonen, T. Van den Bogaert, and J. Wouters, "Reduced-bandwidth and distributed MWF-based noise reduction algorithms for binaural hearing aids," *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 17, no. 1, pp. 38–51, Jan. 2009.

[6] M. Souden, J. Benesty, and S. Affes, "On optimal frequency-domain multichannel linear filtering for noise reduction," *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 18, no. 2, pp. 260–276, Feb. 2010.

[7] B. Cornelis, M. Moonen, and J. Wouters, "Performance analysis of multichannel Wiener filter-based noise reduction in hearing aids under second order statistics estimation errors," *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 19, no. 5, pp. 1368–1381, Jul. 2011.

[8] M. Souden, J. Chen, J. Benesty, and S. Affes, "An integrated solution for online multichannel noise tracking and reduction," *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 19, no. 7, pp. 2159–2169, Sep. 2011.

[9] K. Ngo, A. Spriet, M. Moonen, J. Wouters, and S. H. Jensen, "Incorporating the conditional speech presence probability in multi-channel Wiener filter based noise reduction in hearing aids," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 1, pp. 930625, Jul. 2009.

[10] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 20, no. 4, pp. 1383–1393, May 2012.

[11] P. C. Yong, S. Nordholm, and H. H. Dam, "Noise estimation based on soft decisions and conditional smoothing for speech enhancement," in *Proc. Int. Workshop Acoust. Echo and Noise Control (IWAENC'12)*, Aachen, Germany, Sep. 2012, pp. 4640–4643.

[12] P. C. Yong, S. Nordholm, and H. H. Dam, "Optimization and evaluation of sigmoid function with a priori SNR estimate for real-time speech enhancement," *Speech Communication*, vol. 55, no. 2, pp. 358–376, Feb. 2013.

[13] S. Nordholm, I. Claesson, and M. Dahl, "Adaptive microphone array employing calibration signals: an analytical evaluation," *IEEE Trans. on Speech and Audio Process.*, vol. 7, no. 3, pp. 241–252, May 1999.

[14] N. Grbic and S. Nordholm, "Soft constrained subband beamforming for hands-free speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP'02)*, Orlando, USA, May 2002, vol. 1, pp. 885–888.

[15] H. Q. Dam, S. Y. Low, H. H. Dam, and S. Nordholm, "Space constrained beamforming with source PSD updates," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP'04)*, Montreal, Canada, May 2004, vol. 4, pp. 93–96.

[16] J. E. Greenberg, P. M. Peterson, and P. M. Zurek, "Intelligibility-weighted measures of speech-to-interference ratio and speech system performance.," *J. Acous. Soc. Am.*, vol. 94, no. 5, pp. 3009–3010, Nov. 1993.

[17] P. Loizou, *Speech Enhancement Theory and Practice*, CRC Press, Boca Raton, FL, 2007.

[18] Acoustical Society of America, "ANSI S3.5-1997 American National Standard Methods for calculation of the speech intelligibility index," Jun. 1997.

[19] R. C. Hendriks and R. Martin, "MAP estimators for speech enhancement under normal and rayleigh inverse gaussian distributions," *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 15, no. 3, pp. 918–927, Mar. 2007.